

Автоматизированное обнаружение геологических объектов в изображениях сейсмического поля с применением нейронных сетей глубокого обучения

Ф.В. Краснов

кандидат технических наук

эксперт ООО «Газпромнефть НТЦ»

Адрес: 190000, г. Санкт-Петербург, наб. реки Мойки, д. 75–79, литер Д

E-mail: Krasnov.FV@gazprom-neft.ru

А.В. Буторин

эксперт ООО «Газпромнефть НТЦ»

Адрес: 190000, г. Санкт-Петербург, наб. реки Мойки, д. 75–79, литер Д

E-mail: Butorin.AV@gazpromneft-ntc.ru

А.Н. Ситников

заместитель генерального директора по геологии и разработке месторождений

ООО «Газпромнефть НТЦ»

Адрес: 190000, г. Санкт-Петербург, наб. реки Мойки, д. 75–79, литер Д

E-mail: Sitnikov.AN@gazpromneft-ntc.ru

Аннотация

Бизнес-задача интерпретации данных сейсмических исследований всегда решалась только с помощью высокоуровневых экспертов. Для интерпретации данных сейсмических исследований авторы применили подход на основе компьютерного зрения. Экспертная задача интерпретации сейсмических данных стала частично автоматизированной при помощи методов машинного обучения для классификации изображений, примененных авторами. Для получения набора данных были использованы методы трансформации сейсмических трасс с помощью непрерывного вейвлетного преобразования. В предыдущих работах авторов были созданы методы спектральной декомпозиции, которые также легли в основу данного исследования.

Использование искусственных нейронных сетей глубокого обучения позволило авторам решить задачу классификации изображений. Важно отметить, что требования бизнеса в части режима распространения информации наложили определенные ограничения на объемы используемых вычислительных мощностей и количество размеченных данных. Найденное решение по использованию обученных искусственных нейронных сетей и аугментации изображений помогло успешно справиться с поставленной задачей, несмотря на ограничения.

Полученные результаты позволяют выявлять геологические объекты в сейсмических данных с точностью 90% по метрике F1-score. Это позволило внедрить в существующие бизнес-процессы компании «Газпромнефть НТЦ» автоматизированные процедуры, существенно сокращающие время обработки сейсмических данных.

Социально значимым результатом данного исследования авторы считают найденную возможность «оцифровки» и сохранения знаний лучших экспертов по интерпретации сейсмических данных, а также возможность использования бесконтактной экспертизы для обнаружения геологических объектов в сейсмических данных в периметре группы компаний «Газпром Нефть».

Ключевые слова: машинное обучение с учителем, незначимые наборы данных, спектральная декомпозиция, интерпретация сейсмических данных, компьютерное зрение.

Цитирование: Краснов Ф.В., Буторин А.В., Ситников А.Н. Автоматизированное обнаружение геологических объектов в изображениях сейсмического поля с применением нейронных сетей глубокого обучения // Бизнес-информатика. 2018. № 2 (44). С. 7–16. DOI: 10.17323/1998-0663.2018.2.7.16.

Введение

В настоящем исследовании рассматривается вопрос автоматизации анализа геолого-геофизической информации, с целью оптимизации принимаемых решений. Данное направление активно развивается в стратегии «электронной разработки актива» (ЭРА) (Digital Oil Field, DOF) [1]. Разработка месторождения – это самый продолжительный и затратный этап освоения месторождения, начинающийся с планирования бурения первых скважин и завершающийся через десятки лет деятельности месторождения. Задача инженеров-нефтяников состоит в том, чтобы подобрать наиболее эффективные решения, как при подготовке месторождения, так и на поздних стадиях, чтобы в итоге добывать из недр максимум нефти при наименьших затратах. Стратегия развития цифровых проектов «ЭРА» в сфере разведки и добычи охватывает все основные направления деятельности: геологоразведку, бурение, разработку, добычу и обустройство месторождений.

В области геологоразведки цифровые технологии используются для анализа геологической и геофизической информации о недрах: результатов бурения, скважинных исследований и геофизических работ. Эта информация формирует значительный по объему поток данных, анализ которых позволяет выполнять прогноз геологического строения месторождения и построение геолого-гидродинамических моделей для дальнейшего моделирования разработки и поиска наиболее эффективных технологий добычи.

Данные поверхностных геофизических исследований позволяют получить значительный объем информации о геологической среде на всей территории месторождения. В этой области особенно выделяется сейсморазведка, которая позволяет изучать среду по характеру распространения упругих колебаний в ней. Данный метод требует искусственного возбуждения колебаний в приповерхностной зоне, после чего образовавшиеся упругие волны взаимодействуют со средой и возвращаются обратно, где регистрируются на поверхности.

Результатом последующей обработки зарегистрированных упругих волн является сейсмический образ геологической среды, который формируется сигналами волн, отраженных от упругих контрастов. Одним из вариантов представления данной информации выступает сейсмический куб (для площадных исследований), отображающий в координатах «X–Y–время» амплитуды отраженных сигналов для геологической среды от поверхности до глубин в несколько километров (*рисунк 1*). Дальнейший анализ заключается в изучении распределения амплитуд отраженных волн, которые связаны со свойствами среды, по площади в рамках целевого интервала. Полученная информация интегрируется со скважинными данными и выступает основой оценки геологических свойств изучаемого интервала, то есть формирует геологическую модель месторождения.

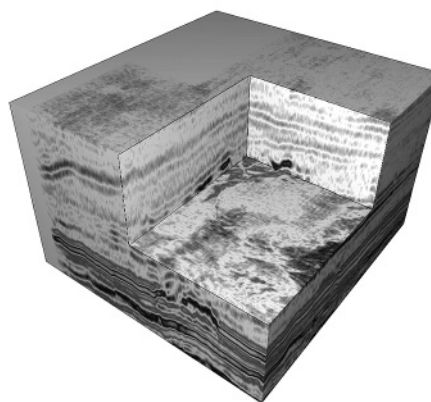


Рис. 1. Пример сейсмического волнового поля (вертикальная ось – время, горизонтальные оси – расстояние, цвет – амплитуда отраженной волны).

Значительная часть современных месторождений находится на поздних стадиях разработки по основным продуктивным пластам, что приводит к необходимости поиска, так называемых «пропущенных интервалов» – перспективных объектов, не затронутых разработкой, которые могут характеризоваться наличием углеводородов в поровом пространстве. Задача поиска подобных объектов предусматривает анализ геологических свойств

всего геологического разреза, не затронутого ранее разработкой. В этой связи широкое распространение получила объектно-ориентированная интерпретация — поиск однозначных геологических объектов, которые выделяются по характеру изменения амплитуды отраженных волн. При этом для получения более информативных результатов исследователи прибегают к различным трансформациям исходного волнового поля с целью выделения той или иной особенности. Примером такой трансформации может выступать спектральная декомпозиция и связанный с ней метод ложноцветового представления, описанный в работе [2] (рисунк 2).

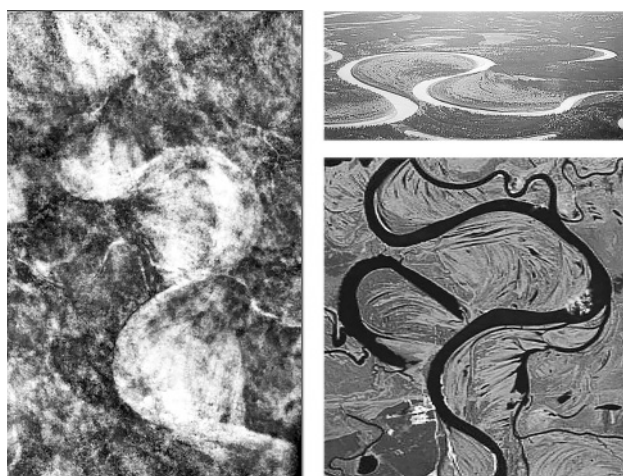


Рис. 2. Сравнение ложноцветового представления сейсмических данных (слева) и современных речных систем (справа)

Процесс поиска «пропущенных интервалов» заключается в анализе распределения аномалий поля внутри интересующего интервала по площади, при этом исследователь вынужден визуально идентифицировать геологические объекты по характерным аномалиям. В данном случае активно используется принцип актуализма, который устанавливает схожесть современных объектов с их аналогами в геологическом прошлом. Таким образом, процесс поиска заключается в решении ряда задач, к числу которых относятся:

- ◆ идентификация целевого интервала;
- ◆ получение площадных оценок аномалий для интервала с детальным шагом по времени/глубине;
- ◆ анализ полученных площадных оценок с целью поиска геологических объектов;
- ◆ совместный анализ выделенных объектов со скважинными данными.

Основной недостаток подобного подхода заключается в значительных трудозатратах, а также в возможном субъективизме при оценке геологических объектов. Исследователю необходимо изучить значительное количество площадных оценок аномалий и выделить перспективные интервалы и участки.

Как отмечалось выше, стратегия «ЭРА» предусматривает автоматизацию процессов, в том числе геологоразведочных, для оптимизации работы на месторождении. В этой связи автоматизация процесса поиска «пропущенных интервалов» является важной бизнес-задачей, как для сокращения временных затрат и получения более объективного результата, не зависящего от конкретного исследователя, так и для получения потенциальных целей для последующего бурения и добычи углеводородов.

В рамках статьи рассматривается решение указанной проблемы с использованием технологий машинного обучения для автоматического поиска и классификации получаемых аномалий. При этом используемый алгоритм должен характеризоваться высокой точностью решения, а также не должен требовать обширной выборки для обучения. В качестве характерной аномалии выбраны речные системы, представленные каналами (руслами), которые представляют значительный интерес с точки зрения наличия коллектора, а также уверенно идентифицируются с точки зрения принципов актуализма.

Таким образом, в статье рассматривается проблема «трансляции» знаний исследователя обучающемуся алгоритму поиска и классификации. Решение данной задачи позволяет повысить скорость принятия решения, а также увеличить глубину исследования, в виду отсутствия процедур, требующих значительных затрат времени.

Искусственные нейронные сети неоднократно применялись для решения широкого круга задач нефтегазовой отрасли. Получившие не так давно новый толчок развития искусственные нейронные сети глубокого обучения пока не так популярны в нефтегазовой отрасли и еще не нашли такого широкого применения, как, например, в медицине и сети Интернет. Основная конструктивная особенность искусственных нейронных сетей глубокого обучения состоит в том, что у них, в отличие от других искусственных нейронных сетей, могут быть десятки слоев с нейронами.

Кроме количества слоев, важным свойством искусственных нейронных сетей глубокого обучения является их архитектура или, иначе говоря, то, из

каких типов нейронов состоят слои и как они соединены между собой. Среди наиболее успешных архитектур искусственных нейронных сетей глубокого обучения можно выделить архитектуру Xception [3]. Точность предсказаний искусственных нейронных сетей глубокого обучения с различными архитектурами представлена в *Таблице 1*.

Таблица 1.

Точность различных архитектур искусственных нейронных сетей глубокого обучения на наборе данных ImageNet [4]

Архитектура	Топ-5 точность	Количество параметров	Количество слоев (глубина)
Xception	0,945	22 910 480	126
VGG16	0,901	138 357 544	23
VGG19	0,910	143 667 240	26
ResNet50	0,929	25 636 712	168
InceptionResNetV2	0,953	55 873 736	572

Обучение искусственных нейронных сетей – сложная инженерно-техническая задача. Для распознавания изображений в качестве обучающей выборки часто используют базу размеченных изображений ImageNet. В базе изображений ImageNet реализована иерархическая классификация изображений, в которой узел (существительное) может содержать сотни и тысячи изображений. В среднем сейчас на каждый узел ImageNet приходится по пятьсот изображений. Например, узел может быть «кошки» или «люди». Обычно обучение производится на каком-то подмножестве узлов. Однако среди узлов ImageNet нет геологических объектов, в которых мы заинтересованы для данного исследования.

В архитектуре Xception более 22 миллионов параметров, значения которых нужно получить в результате процесса обучения. Успешное обучение Xception, описанное в работе [5], было осуществлено на высокопроизводительном вычислительном кластере, которого у авторов данного исследования не было в наличии.

Кроме технических условий, предстояло найти метод формализации и «оцифровки» человеческой экспертизы в отношении процесса идентификации геологических объектов в сейсмических данных.

С учетом всех вышеизложенных условий авторами была выдвинута следующая исследовательская гипотеза: «Существует алгоритм, использующий

принципы машинного обучения, позволяющий осуществлять высокоточную идентификацию геологических объектов в сейсмических данных за время, сравнимое с затрачиваемым на идентификацию экспертами, и использующий для обучения небольшой по объему набор данных».

Для проверки гипотезы было проведено исследование, состоящее из описания использованных и разработанных авторами методов, цифрового эксперимента и заключения.

1. Методы

Основой исследования является решение задачи классификации изображений с помощью обучения с учителем (supervised learning) на малых наборах данных. Схематически подход, предложенный авторами, изображен в виде каркаса исследования (*рисунок 3*).



Рис. 3. Методический каркас исследования

Далее каждый из элементов исследовательского каркаса рассмотрен подробнее с точки зрения методики обработки данных и построения алгоритмов.

1.1. Сейсмические кубы

С точки зрения данных, сейсмические кубы представляют собой файлы в формате SEG-Y. SEG-Y – это открытый формат данных, содержащий акустические трассы. Размер SEG-Y файлов может колебаться от 1 ГБ до 100 ГБ. В одном таком файле содержится информация об определенной площади месторождения – это может быть площадь разме-

ром в десятки квадратных километров. Количество трасс в одном кубе может быть в среднем несколько миллионов. Каждая трасса обладает шагом дискретизации по времени в 1–4 мс. Таким образом, общая трехмерная сетка куба может быть представлена в среднем размерностью (3000, 3000, 2000). После конвертации каждый сейсмический куб представляется в формате Python NumPy.

1.2. Дизайн признаков

Поиск геологических объектов ведется в определенном диапазоне сейсмического времени (ось Z). Для задания верхней и нижней границ этого диапазона используются секущие поверхности, называемые отражающими горизонтами (ОГ). Определение оптимального ОГ — это отдельная оптимизационная задача, поэтому для ее выполнения выбранный диапазон делится на несколько конформных ОГ. В каком из выделенных ОГ проявит себя геологический объект на данном этапе, точно не известно.

Для того чтобы перейти из пространства координат и глубин в пространство координат и частот для выбранного диапазона, ограниченного ОГ, выполняется непрерывное вейвлет-преобразование (continuous wavelet transformation, CWT).

После этого каждое сечение частотного куба отражающим горизонтом представляет двумерную карту амплитуд частот. Для визуализации таких карт используется механизм RGB смешивания (RGB blending), рассмотренный в работе [6]. Количество получаемых RGB карт можно оценить по формуле:

$$NumOfRGBmaps = ReflHorDim \cdot C_{FrequencyDim}^{RGBdim}, \quad (1)$$

где $NumOfRGBmaps$ — количество RGB карт;

$ReflHorDim$ — количество конформных горизонтов в выбранном диапазоне;

$C_{FrequencyDim}^{RGBdim}$ — количество сочетаний из количества выбранных частот ($FrequencyDim$) по количеству цветов ($ColorDim$) для рассматриваемого случая $ColorDim = 3(RGB)$.

Полученные двумерные RGB карты являются исходными данными для поиска геологических объектов.

1.3. Создание набора данных

Основная задача создания набора данных состоит в том, чтобы подготовить полученные RGB карты для процесса машинного обучения. RGB карта

может содержать или не содержать геологические объекты. Для разметки набора данных необходимо вручную отнести данную RGB карту к определенному классу.

На основании размеченного набора данных возможно решить задачу бинарной классификации. Отметим, что размерность RGB карты равна рассматриваемой площади месторождения. Как упоминалось ранее, это в среднем (3000, 3000). Для эффективного использования алгоритмов компьютерного зрения RGB карты было принято решение разбить ее на изображения размером (224, 224). Таким образом, для каждой карты получилось в среднем 169 изображений (мини RGB карт).

Полная разметка полученного набора данных — это около 16 миллионов мини RGB карт. Поэтому было принято решение разметить небольшое подмножество мини RGB карт, состоящее из 50 мини RGB карт с каналами и 50 мини RGB карт без каналов.

1.4. Обучение глубокой нейронной сети

Мы рассмотрели следующие архитектуры глубоких нейронных сетей:

✦ VGG16 — сеть Visual Geometry Group из университета Оксфорда для распознавания объектов на изображениях. Состоит из 16 слоев [7];

✦ Inception v3 — нейронная сеть компании Google для распознавания объектов на изображениях. Состоит из 159 слоев [8];

✦ ResNet50 — нейронная сеть компании Microsoft, использующая остаточное обучение (residual learning). Применяется для распознавания объектов на изображениях. Состоит из 168 слоев [9];

✦ Xception — модификация сети Inception, содержащая 126 слоев [3].

Руководствуясь сравнениями точности вышеперечисленных архитектур, нами была выбрана наиболее современная архитектура Xception.

«Нет достаточного количества процессоров и размеченных данных» — так образно можно назвать ситуацию, в которой авторы проводили свое исследование. Такая ситуация является гораздо более распространенной и многие исследовательские коллективы сетуют на нее в частных коммуникациях, как на непреодолимый барьер. Однако авторам данного исследования удалось ее преодолеть.

Обученная на наборе данных ImageNet искусственная нейронная сеть с архитектурой Xception предназначена для классификации 1000 классов бытовых объектов из ImageNet. Геологических объектов в наборе данных ImageNet нет. Поэтому авторы данного исследования воспользовались методикой переноса знаний. Суть методики переноса знаний состоит в том, чтобы использовать только часть обученных слоев, разработать собственные слои и доучить искусственную нейронную сеть с архитектурой Xception на данных из необходимого проблемного домена, в данном случае – на изображениях мини RGB карт.

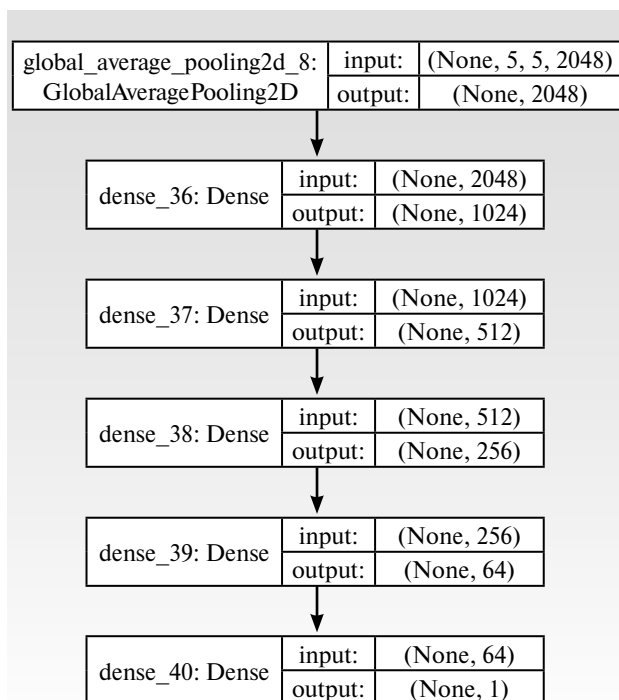


Рис. 4. Добавленные слои (№№36–40)

Авторами были добавлены один «pooling» и пять «full connected» слоев (рисунк 4) с размерностями (1024, 512, 256, 64, 1).

Таким образом, была создана возможность обучения слоев, отображенных на рисунке 4.

Для искусственного увеличения набора данных авторами применена методика аугментирования изображений, усовершенствованная на основе исследования [10]. К каждому размеченному изображению был применен определенный набор трансформаций. В результате количество различных изображений в наборе данных было увеличено в 12 раз. На рисунке 5 приведен пример аугментирования одной мини RGB карты.

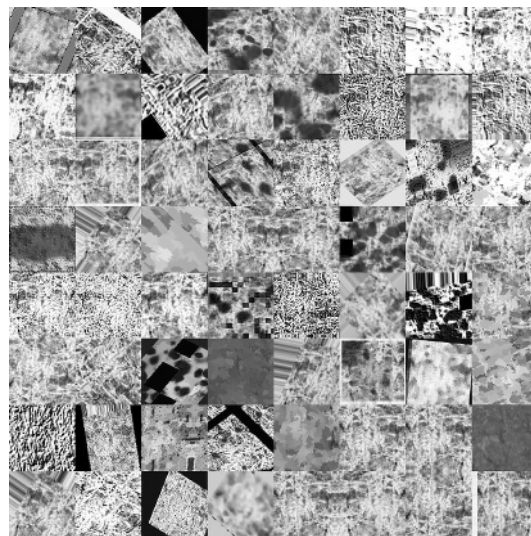


Рис. 5. Пример аугментирования одной мини RGB карты

2. Результаты эксперимента

Для обучения модели были использованы следующие оптимизационные алгоритмы:

- ◆ Stochastic Gradient Descent (SGD) optimizer;
- ◆ RMSProp optimizer;
- ◆ Adam optimizer.

В качестве функции потерь в процессе оптимизации использована функция перекрестной энтропии:

$$L_{\log}(y, p) = -(y \cdot \log(p) + (1 - y) \cdot \log(1 - p)), \quad (2)$$

где истинный класс $y \in \{0, 1\}$, а p – вероятность того, что $y = 1$.

Обучение модели производилось в двух режимах:

Режим обучения I. Обучение производится только по добавленным авторами слоям. Параметры, полученные при обучении модели на данных ImageNet не изменяются. Кривые обучения для различных оптимизаторов с уже подобранными параметрами (*learning rate* и другими специфичными для каждого оптимизатора параметрами) представлены на рисунке 6.

Режим обучения II. Производится обучение всех параметров. Параметры, полученные при обучении модели на данных ImageNet, используются как начальные. Кривые обучения для различных оптимизаторов с уже подобранными параметрами (*learning rate* и другими специфичными для каждого оптимизатора параметрами) представлены на рисунке 8.

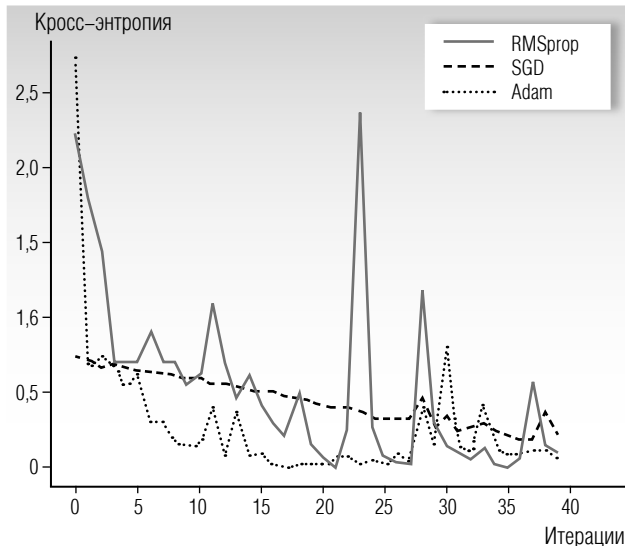


Рис. 6. Кривые обучения для различных оптимизаторов в Режиме I

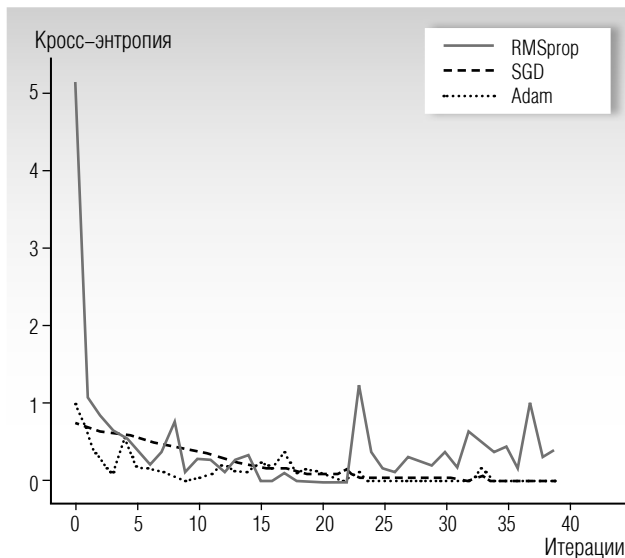


Рис. 7. Кривые обучения для различных оптимизаторов в Режиме II

В итоге авторы остановились на оптимизационном алгоритме Adam [11] в Режиме II с параметрами оптимизатора $learning\ rate = 3,16e - 4$; $\beta_1 = 0,9$; $\beta_2 = 0,999$. Процесс обучения занимал в среднем около 18 часов. Отметим, что в Режиме I процесс обучения занимал около часа. Такое различие в скорости обучения объясняется различием в количестве параметров. Оптимизатор Adam показал более быструю сходимость и производительность.

Для проверки качества обучения модели использовалась метрика *Accuracy*. На рисунке 8 представлены значения метрики *Accuracy* в зависимости от шага обучения для различных параметров $learning\ rate$ выбранного оптимизатора Adam в Режиме II.

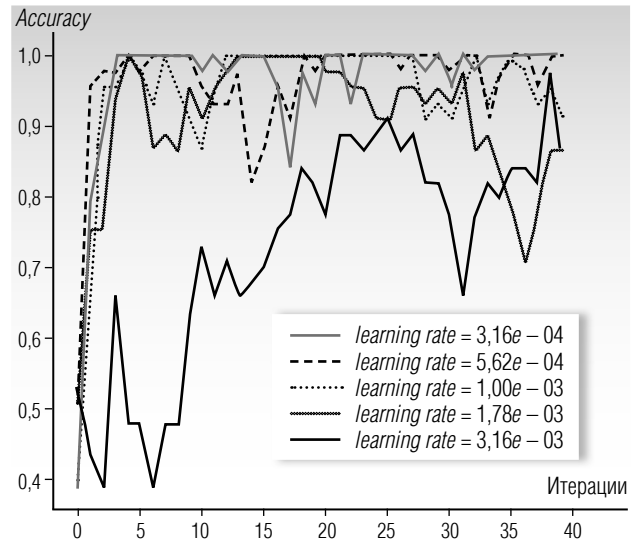


Рис. 8. Значения метрики *Accuracy* в зависимости от шага обучения для различных параметров $learning\ rate$ выбранного оптимизатора Adam в Режиме II

Для выбора порога вероятности принятия положительного решения по отнесению к классу был сделан анализ метрик *Precision u Recall*. В условиях поставленной задачи нас в большей степени интересовало обнаружение каналов на RGB картах. Наличие небольшого количества случаев определения *False Positive* не является критичным. Итоговые значения метрик точности представлены в Таблице 2.

Таблица 2.

Отчет о классификации

	Precision	Recall	F1-score	Support
Отсутствие каналов	0,89	0,92	0,91	26
Наличие каналов	0,91	0,88	0,89	24
Среднее / Всего	0,90	0,90	0,90	50

На основе обученной модели была произведена классификация изображений RGB карт для одного месторождения. На рисунке 9 приведены две карты RGB, отнесенные к разным классам на основании решения модели.

Заключение

Современный уровень открытости сообщества по развитию глубоких нейронных сетей позволяет небольшим исследовательским организациям решать свои производственные задачи с помощью наиболее передовых алгоритмов. Публикация в открытом доступе не только программного кода, но и параметров обученных глубоких нейронных сетей выводит

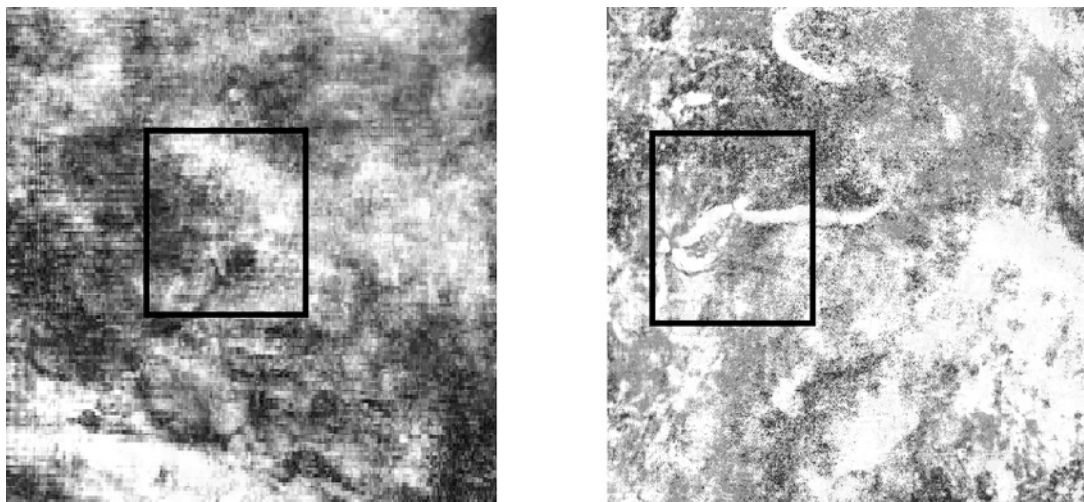


Рис. 9. Две RGB карты: левая классифицирована как имеющая каналы, правая каналов не имеет. Квадрат 224, 224 обозначает область, на основании которой принималось решение о классификации

на новый уровень возможности классификации геологических изображений.

В данном исследовании авторами получены результаты, по метрике *Accuracy* сравнимые с заявленными для данной архитектуры глубокой нейронной сети. Среднее значение метрики *F1-score* составило 90%.

Авторами успешно использованы разработанные ими в предыдущих работах методы спектральной декомпозиции. В настоящем исследовании эти методы получили применение в создании набора данных.

Методика аугментации изображений показала себя гибким и универсальным инструментом расширения набора данных. Данный подход может быть использован во многих областях, где возможности ручного разбиения набора данных на классы для обучения ограничены.

Решенная авторами задача классификации геологических объектов обладает новизной и имеет прак-

тическое применение в научно-исследовательских организациях нефтегазовой отрасли.

Полученные авторами автоматизированные процедуры позволяют увеличить точность определения геологических объектов, что особенно важно в условиях трудноизвлекаемых запасов.

Используемые в настоящее время методы интерпретации ложноцветовых изображений сейсмических полей основаны на эвристических подходах. Новизна данного исследования состоит в том, чтобы сохранить существующие эвристические подходы с помощью обучения глубоких нейронных сетей.

Методы выявления геологических объектов с помощью глубоких нейронных сетей являются примером подхода «обучения с учителем», который позволяет продвинуться в сложном процессе перехода к концепции «цифрового месторождения» путем аккумуляции знаний экспертов для многократного применения. ■

Литература

1. Digital information systems creation for optimization of complex geotechnical jobs programs formation process for JSC Gazprom Neft oilfields / A. Sitnikov [et al.] // Proceedings of SPE Russian Petroleum Technology Conference, Moscow, Russia, 26–28 October 2015. Vol. 1. P. 642–662.
2. Butorin A.V., Krasnov F.V. Approaches to the analysis of spectral decomposition for the purpose of detailed geological interpretation // Proceedings of SPE Russian Petroleum Technology Conference, Moscow, Russia, 24–26 October 2016. Vol. 1. P. 275–289.
3. Chollet F. Xception: Deep learning with depthwise separable convolutions // arXiv:1610.02357v3 [cs. CV]. 2017. [Электронный ресурс]: <https://arxiv.org/pdf/1610.02357.pdf> (дата обращения 01.02.2018).
4. He K., Zhang X., Ren S., Sun J. Identity mappings in deep residual networks // Proceedings of the 14th European Conference on Computer Vision (ECCV 2016). Amsterdam, The Netherlands, 11–14 October 2016. Part IV. Lecture Notes in Computer Science. Vol. 9908. Springer, Cham, 2016. P. 630–645.
5. TensorFlow: Large-scale machine learning on heterogeneous distributed systems / M. Abadi [et al.] // Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16). Savannah, GA, USA, 2–4 November 2016. P. 265–283.
6. Li Y.-F., Cheng J.-Y., Zhu S.-J., Wang C. Seismic multi-attribute analysis based on RGB color blending technology // Journal of China Coal Society. 2009. No. 11. P. 018–034.

7. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // arXiv:1409.1556v6 [cs. CV]. 2014. [Электронный ресурс]: <https://arxiv.org/pdf/1409.1556.pdf> (дата обращения 01.02.2018).
8. Rethinking the inception architecture for computer vision / C. Szegedy [et al.] // Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, NV, USA, 26 June – 1 July 2016. P. 2818–2826.
9. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, NV, USA, 26 June – 1 July 2016. P. 770–778.
10. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. 2012. P. 1097–1105.
11. Kingma D., Ba J. Adam: A method for stochastic optimization // arXiv:1412.6980v9 [cs. LG]. 2017. [Электронный ресурс]: <https://arxiv.org/pdf/1412.6980.pdf> (дата обращения 01.02.2018).

Automatic detection of channels in seismic images via deep learning neural networks

Fedor V. Krasnov

Expert, Science & Technology Centre

Gazprom Neft PJSC

Address: 75–79 liter D, Moika River Embankment, Saint Petersburg, 190000, Russian Federation

E-mail: krasnov.fv@gazprom-neft.ru

Alexander V. Butorin

Expert, Science & Technology Centre

Gazprom Neft PJSC

Address: 75–79 liter D, Moika River Embankment, Saint Petersburg, 190000, Russian Federation

E-mail: Butorin.AV@gazpromneft-ntc.ru

Alexander N. Sitnikov

Deputy Chief Executive Officer for Geology and Field Development, Science & Technology Centre

Gazprom Neft PJSC

Address: 75–79 liter D, Moika River Embankment, Saint Petersburg, 190000, Russian Federation

E-mail: Sitnikov.AN@gazpromneft-ntc.ru

Abstract

The business goal of interpreting seismic data has always been addressed by the high-level experts engaged. The authors applied a computer vision approach to interpret seismic data. The expert task of interpreting seismic data has become partially automated via machine learning techniques utilized to classify the images used by the authors. The methods of transformation of seismic traces through spectral decomposition were used to obtain the data set. In the previous works of the authors, methods of spectral decomposition via continuous wavelet transformation were created, and this also laid the foundation of this study.

Use of artificial neural networks of deep learning has enabled the authors to meet the goal of image classification. In this regard, it is important to note that the business policy related to information dissemination imposed certain limitations on the computing capacity used and the number of the data labeled. The solution found for the use of trained artificial neural networks and image augmentation helped us to successfully cope with the goal, in spite of the limitations.

The results obtained allow us to identify geological units with a test accuracy of 90% rendering to the F1-score measure. This has enabled the Scientific and Technical Center of Gazprom Neft to implement automated procedures in the existing business processes in order to significantly reduce the time needed to process seismic data.

The authors consider the possibility of “digitizing” and preserving the knowledge of the highest-level experts in interpreting seismic data, as well as the possibility of using contactless examination to locate geological units in the seismic data within the Gazprom Neft group of companies to be a socially efficient outcome of this study.

Key words: supervised machine learning, small datasets, spectral decomposition, seismic, geological units, computer vision.

Citation: Krasnov F.V., Butorin A.V., Sitnikov A.N. (2018) Automatic detection of channels in seismic images via deep learning neural networks. *Business Informatics*, no. 2 (44), pp. 7–16. DOI: 10.17323/1998-0663.2018.2.7.16.

References

1. Sitnikov A., Pustovskikh A., Asmandiyarov R., Gilmanov R., Sheremeev A., Zulkarniev R. (2015) Digital information systems creation for optimization of complex geotechnical jobs programs formation process for JSC Gazprom Neft oilfields. *Proceedings of SPE Russian Petroleum Technology Conference, Moscow, Russia, 26–28 October 2015*, vol. 1, pp. 642–662.
2. Butorin A.V., Krasnov F.V. (2016) Approaches to the analysis of spectral decomposition for the purpose of detailed geological interpretation. *Proceedings of SPE Russian Petroleum Technology Conference, Moscow, Russia, 24–26 October 2016*, vol. 1, pp. 275–289.
3. Chollet F. (2017) *Xception: Deep learning with depthwise separable convolutions*. arXiv:1610.02357v3 [cs. CV]. Available at: <https://arxiv.org/pdf/1610.02357.pdf> (accessed 01 February 2018).
4. He K., Zhang X., Ren S., Sun J. (2016) Identity mappings in deep residual networks. *Proceedings of the 14th European Conference on Computer Vision (ECCV 2016). Amsterdam, The Netherlands, 11–14 October 2016*, part IV, *Lecture Notes in Computer Science*, vol. 9908. Springer, Cham, pp. 630–645.
5. Abadi M., Agarwal A., Barham P., et al. (2016) TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16). Savannah, GA, USA, 2–4 November 2016*, pp. 265–283.
6. Li Y.-F., Cheng J.-Y., Zhu S.-J., Wang C. (2009) Seismic multi-attribute analysis based on RGB color blending technology. *Journal of China Coal Society*, no. 11, pp. 018–034.
7. Simonyan K., Zisserman A. (2014) *Very deep convolutional networks for large-scale image recognition*. arXiv:1409.1556v6 [cs. CV]. Available at: <https://arxiv.org/pdf/1409.1556.pdf> (accessed 01 February 2018).
8. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. (2016) Rethinking the inception architecture for computer vision. *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, NV, USA, 26 June – 1 July 2016*, pp. 2818–2826.
9. He K., Zhang X., Ren S., Sun J. (2016) Deep residual learning for image recognition. *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, NV, USA, 26 June – 1 July 2016*, pp. 770–778.
10. Krizhevsky A., Sutskever I., Hinton G.E. (2012) ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pp. 1097–1105.
11. Kingma D., Ba J. (2017) *Adam: A method for stochastic optimization*. arXiv:1412.6980v9 [cs. LG]. Available at: <https://arxiv.org/pdf/1412.6980.pdf> (accessed 01 February 2018).